Bailando: 3D Dance Generation by Actor-Critic GPT with Choreographic Memory

Abstract

Driving 3D characters to dance following a piece of music is highly challenging due to the spatial constraints applied to poses by choreography norms. In addition, the generated dance sequence also needs to maintain temporal coherency with different music genres. To tackle these challenges, we propose a novel music-to-dance framework, Bailando, with two powerful components: 1) a choreographic memory that learns to summarize meaningful dancing units from 3D pose sequence to a quantized codebook, 2) an actor-critic Generative Pre-trained Transformer (GPT) that composes these units to a fluent dance coherent to the music. With the learned choreographic memory, dance generation is realized on the quantized units that meet high choreography standards, such that the generated dancing sequences are confined within the spatial constraints. To achieve synchronized alignment between diverse motion tempos and music beats, we introduce an actor-critic-based reinforcement learning scheme to the GPT with a newly-designed beat-align reward function. Extensive experiments on the standard benchmark demonstrate that our proposed framework achieves state-of-the-art performance both qualitatively and quantitatively. Notably, the learned choreographic memory is shown to discover human-interpretable dancing-style poses in an unsupervised manner. Code and video demo are available at https://github.com/lisiyao21/Bailando/